

E-RNAi: a web application for the multi-species design of RNAi reagents—2010 update

Thomas Horn and Michael Boutros*

German Cancer Research Center (DKFZ), Division Signaling and Functional Genomics and University of Heidelberg, Department of Cell and Molecular Biology, Faculty of Medicine Mannheim, Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany

Received January 30, 2010; Revised April 5, 2010; Accepted April 17, 2010

ABSTRACT

The design of RNA interference (RNAi) reagents is an essential step for performing loss-of-function studies in many experimental systems. The availability of sequenced and annotated genomes greatly facilitates RNAi experiments in an increasing number of organisms that were previously not genetically tractable. The E-RNAi web-service, accessible at <http://www.e-rnai.org/>, provides a computational resource for the optimized design and evaluation of RNAi reagents. The 2010 update of E-RNAi now covers 12 genomes, including *Drosophila*, *Caenorhabditis elegans*, human, emerging model organisms such as *Schmidtea mediterranea* and *Acyrtosiphon pisum*, as well as the medically relevant vectors *Anopheles gambiae* and *Aedes aegypti*. The web service calculates RNAi reagents based on the input of target sequences, sequence identifiers or by visual selection of target regions through a genome browser interface. It identifies optimized RNAi target-sites by ranking sequences according to their predicted specificity, efficiency and complexity. E-RNAi also facilitates the design of secondary RNAi reagents for validation experiments, evaluation of pooled siRNA reagents and batch design. Results are presented online, as a downloadable HTML report and as tab-delimited files.

INTRODUCTION

Functional genomic studies by RNA interference (RNAi) are now widely used to determine gene functions in a variety of cell-based and *in vivo* model systems (1). RNAi is triggered by exogenous double-stranded (ds) RNAs causing the degradation of endogenous messenger RNAs (2). The conservation of RNAi from plants to

humans (3) allows its application in a broad spectrum of model organisms for which genome sequences and gene annotations are available.

RNAi experiments in invertebrate model systems mainly use long dsRNAs (100–700-bp long) (4) that are intracellularly cleaved into 19–23-bp long siRNAs by Dicer (5). In vertebrate cells, siRNAs (21-nt duplexes) are used to trigger mRNA knock-downs (6). Over the past years, many sequence properties affecting the silencing specificity and efficiency of siRNAs have been identified. Unintended effects can for example be caused by the overall homology of siRNAs to ‘off-target’ mRNAs (7), miRNA-like seed regions contained within siRNAs (8,9), and low-complexity regions (e.g. CA[ACGT] or CAN repeats) (10). Sequences that contain such motifs should be avoided in the design of long dsRNAs and siRNAs. Key parameters that influence the efficiency of introduced siRNAs were discovered in optimization studies in mammalian cells, including differential end stability of siRNAs, GC content, unstable or no secondary structures of the siRNAs and their targets as well as several base preferences (11–13).

Considering the potential pitfalls of RNAi experiments resulting e.g. from unspecific effects or poor knock-down efficiency by RNAi reagents (14), computational tools are necessary to enable the optimized design and evaluation of long dsRNAs and siRNAs in a user-friendly manner. A limited number of web-based tools are available for the design of long dsRNAs (e.g. DEQOR (15), SnapDragon at <http://www.flyrnai.org/>). The collection of online services available for the design of siRNAs is diverse including commercial design tools [e.g. siDESIGN Center (Dharmacon, ThermoScientific), BioPredsi (16), SVM siRNA design tool (13)] and tools that are focused on certain siRNA properties such as the prediction of target-site accessibility [RNAxs (17)] or siRNA efficiency [siR (12)]. However, most tools are dedicated to a certain model organism, have not implemented batch queries or evaluation of pre-existing RNAi reagents.

Here, we present an updated version of the E-RNAi web service (18) that has been improved by the

*To whom correspondence should be addressed. Tel: +49 6221 421950; Fax: +49 6221 421959; Email: m.boutros@dkfz.de

implementation of new design options and prediction of RNAi reagents for a spectrum of model organisms. E-RNAi now covers 12 model organism that are tractable by RNAi (Table 1). The flexible data structure behind E-RNAi allows for the addition of further organisms. Besides long dsRNAs, siRNA reagents can be designed or evaluated, including a feature for the summarized analysis of siRNA pools. In addition, the user can visually select target regions for the *de novo* design of RNAi reagents from a genome browser interface [GBrowse (19)]. Once a target region is selected, different options such as specificity and efficiency parameters or primer-design settings can be adjusted. Additional quality measurements of RNAi reagents are available, including the evaluation of long dsRNA for low-complexity regions (e.g. CAN repeats), imperfect homologies to other target regions and recent improvements in efficiency calculation methods.

In addition, E-RNAi allows the straightforward design of independent RNAi reagents e.g. by submitting sequences to be excluded. The new version of E-RNAi incorporates extended reports, including the ability to export results as generic feature format (GFF) and annotation file format (AFF) to facilitate their visualization in a genome browser. A full report that contains all results can be downloaded as archive file and locally displayed in any web browser. In addition, the implementation of new alignment tools [Bowtie (20), BLAT (21)] significantly improves the speed of E-RNAi. Batch queries with up to 50 designs can be analyzed in one query.

WEB APPLICATION

The E-RNAi web service automates the steps required for the design and evaluation of RNAi reagents. Until now, queries for 12 genomes have been implemented which can be extended upon the request of users (Table 1). To design or evaluate RNAi reagents, E-RNAi performs several tasks, including: (i) defining target sequences and design options based on user-input, (ii) filtering of

low-complexity regions from input sequences, (iii) *in silico* dicing of input sequences into all possible siRNAs, (iv) evaluation of predicted specificity and (v) predicted efficiency for each siRNA, (vi) design of optimal PCR primer pairs for all regions meeting criteria (ii–v) (for long dsRNAs only), (vii) sorting amplicons or siRNAs for specificity and efficiency, (viii) evaluation of reagents' overall homology to other transcripts, (iv) mapping of reagents to the genome and (x) generation of HTML and tab-delimited reports. In case the user-defined criteria were not met (steps ii–vi) a re-design method (if enabled) will relax criteria and re-start the design process (at step vi). The general workflow and database structure of E-RNAi are shown in Figure 1.

Sequence selection and design/evaluation options

E-RNAi offers three run options: 'ID or sequence input', 'GBrowse input' and 'evaluation of RNAi reagents'. All are accessible from the entry webpage. The first two options allow the selection of target sequences for the *de novo* design of reagents. The third option enables the evaluation of existing RNAi reagents. All input options are available for long dsRNAs and siRNAs, and for all implemented genomes (red panel, Figure 1).

E-RNAi accepts sequences in FASTA or raw formats. The sequences are first mapped to the genome by a local Blat server (21) and returned to the user for target region selection. In case gene identifiers were queried E-RNAi lists sequences for the gene, annotated isoforms and exons (see Table 1 for examples). This also facilitates the selection of exons common to all transcripts (Figure 2A). A more detailed and interactive way to select target sequences is offered by the option to select target regions through a genome browser (GBrowse) interface (19). E-RNAi is connected to a local GBrowse installation where specific regions can be searched by gene-identifier or by absolute genome coordinates. Individual sequences can then be visually selected and submitted to E-RNAi as design templates (Figure 2B).

Table 1. Organisms available in E-RNAi version 3

Organism	Database	Available query identifiers
<i>Drosophila melanogaster</i>	FlyBase	FBgn, CG identifiers and gene symbols (e.g. CG11992, FBgn0014018 or rel)
<i>Acyrtosiphon pisum</i>	AphidBase	ACYPI, XM and LOC identifiers (from 'official gene consensus set', e.g. ACYPI006699, XM_001944230 or LOC100165774)
<i>Homo sapiens</i>	NCBI RefSeq	Gene names, Entrez gene identifiers and RefSeq identifiers (e.g. RPS11, 6205 or NM_001015.3)
<i>Schmidtea mediterranea</i>	SmedGD	MAKER identifiers (mk4 from SmedGD, e.g. mk4.023206.00)
<i>Saccharomyces cerevisiae</i>	EnsemblFungi	Ensembl gene identifiers (e.g. YHR055C)
<i>Schizosaccharomyces pombe</i>	EnsemblFungi	Ensembl gene identifiers (e.g. SPMIT.05)
<i>Caenorhabditis elegans</i>	WormBase	WBGene identifiers, sequence names and gene names (e.g. WBGene00006763, JC8.10 or unc-26)
<i>Anopheles gambiae</i>	VectorBase	AGAP gene identifiers (e.g. AGAP000009)
<i>Aedes aegypti</i>	VectorBase	AAEL gene identifiers (e.g. AAEL000068)
<i>Mus musculus</i>	NCBI RefSeq	Gene names, Entrez gene identifiers and RefSeq identifiers (e.g. Axin1, 12005 or NM_009733.2)
<i>Apis mellifera</i>	BeeBase	GB gene identifiers (e.g. GB15421)
<i>Tribolium castaneum</i>	BeetleBase	TC and GLEAN-prediction gene identifiers (e.g. TC004684 or GLEAN_04684)

Workflow E-RNAi (version 3)

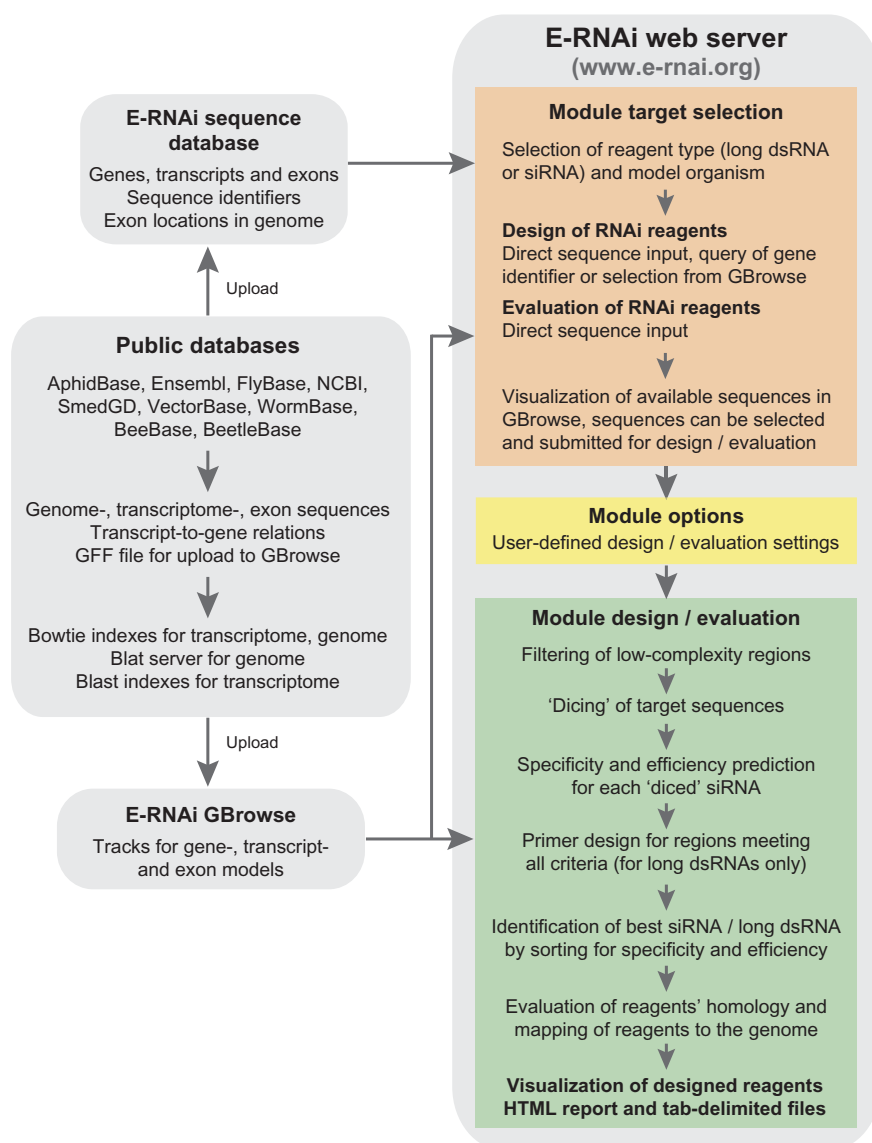


Figure 1. Structure of the E-RNAi web service and database back end. The left side (grey panels) shows the sequences and mapping information required from public databases to populate the E-RNAi sequence- and GBrowse-databases and to generate the indices needed for Bowtie, Blat and Blast alignment tools. The right side shows the workflow of the E-RNAi web server from the selection of sequences and adjustment of options (red and yellow panels) over the processing steps performed during the design of reagents until the generation of a HTML report (green panel). Further details are explained in the text.

In case an organism is not available in E-RNAi, the web service can be run without the definition of an organism ('No off-target database'). The design and evaluation of RNAi reagents is then solely based on efficiency and low-complexity predictions (as well as on primer designs).

After selection of target regions, options for the design or evaluation of reagents can be chosen (yellow panel, Figure 1). All options available in E-RNAi are summarized and described in Table 2.

Assessing the predicted specificity of RNAi reagents

dsRNA sequences have been shown to exert off-target effects via regions of low sequence-complexity (e.g. CAN

repeats (10)) or homology of contained siRNAs to unintended targeted transcripts (7). E-RNAi avoids low-complexity regions by applying the mdust filter program (<http://compbio.dfci.harvard.edu/tgi/software/>) on the queried sequences. Mdust is used with default parameters that will exclude e.g. most simple nucleotide repeats and poly-triplet sequences. Additionally, sequences are filtered for stretches containing more than six contiguous CAN repeats. E-RNAi also 'dices' all queried sequences in siRNAs of a user-defined length (16–28 bp) using a 1 bp shifting window. These are mapped to the transcriptome using Bowtie (20) to assess their specificity. siRNAs that target multiple independent transcripts are flagged. The user can also provide a FASTA sequence file (e.g.

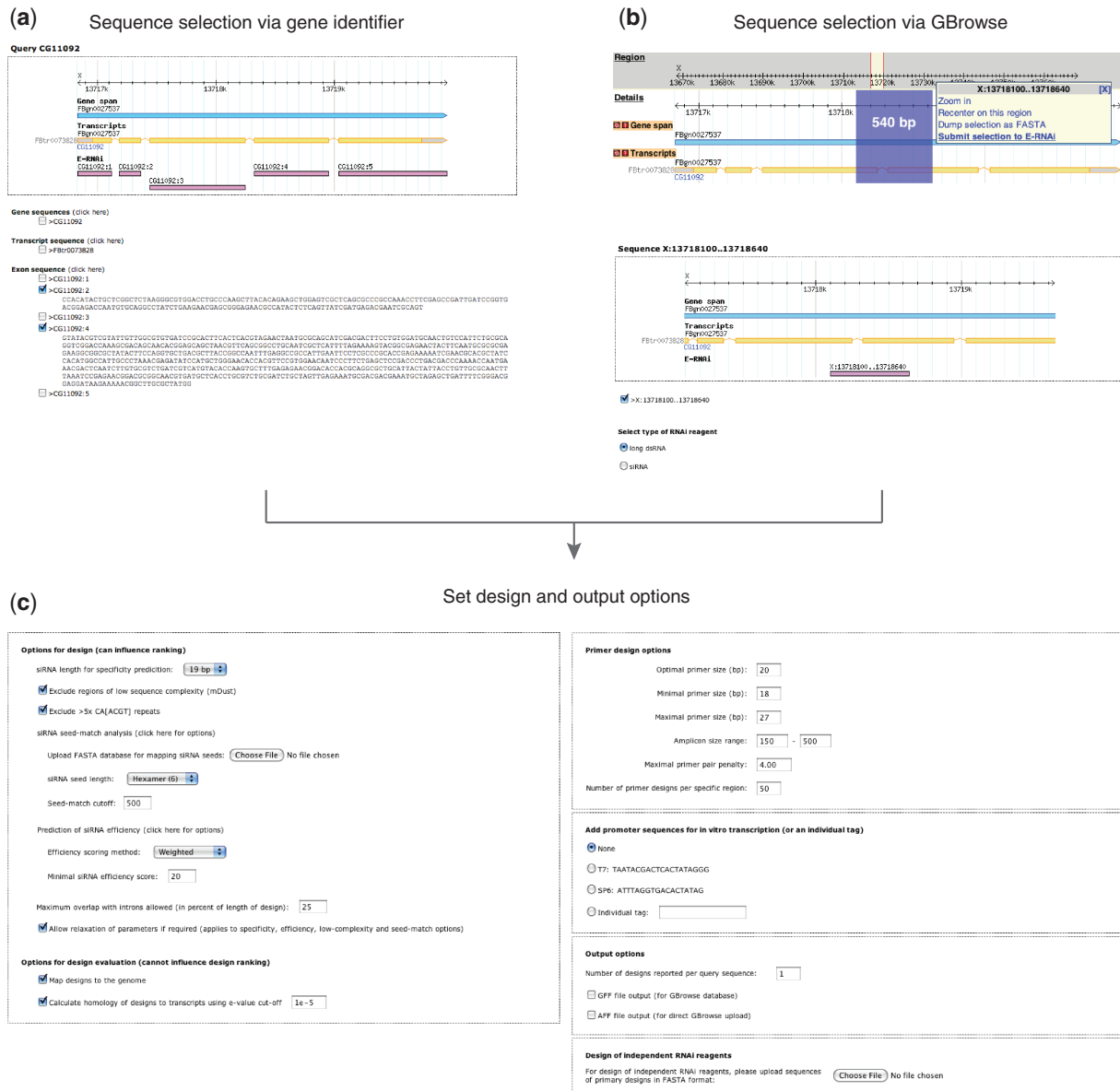


Figure 2. Sequence selection and design options. **(a)** After querying E-RNAi with gene identifiers, gene sequences, transcript sequences or exon sequences can be selected for design. **(b)** Individual sequences can be selected as design templates visually from GBrowse (upper panel). After sequence submission to E-RNAi, the type of reagent (long dsRNA or siRNA) can be selected (lower panel). **(c)** Options available for the design of long dsRNAs include settings for the prediction of specificity, efficiency and low-complexity regions (left panel) as well as primer design settings (upper right panel). In addition different output options, such as the number of designs per query sequence and different report formats, can be selected. The sequence of a previous design can also be uploaded in FASTA format to exclude it from a new (independent) design (lower right panel). Details for certain options are provided in the text and listed in Table 2.

containing 3'-UTR sequences) that is used by E-RNAi to calculate the number of siRNA seed-matches. siRNAs with a seed-match frequency above a user-defined cutoff are flagged.

Additionally, designed reagents are evaluated for partial homologies to unintended targets by Blast (22) searches against the transcriptome with a user-defined E-value cutoff.

Assessing predicted RNAi efficiency

E-RNAi implements two scoring methods to predict the efficiency of siRNAs. The so-called 'rational' (11) and the 'weighted' (12) scoring methods were both developed from

optimization studies using siRNA experiments in human cells. All siRNA features that are included in the score are listed in detail in Supplementary Table S1. Since the scoring methods result in different ranges, we normalized scores to a range between 0 and 100. According to the normalized score, Reynolds *et al.* (11) considered siRNAs with scores ≥ 66.7 as efficient silencers and Shah *et al.* (12) found siRNAs with scores ≥ 63 to be the potent. One of the main differences between both methods is that Reynolds *et al.* (11) suggests the absence of 'internal repeats' (the melting temperature of potential hairpins should be below 20°C) a significant feature, whereas Shah *et al.* did not. E-RNAi assesses potential secondary

Table 2. Long dsRNA / siRNA design options in E-RNAi version 3

Parameter	Option	Reagent	Description ^a	Default
siRNA length for specificity prediction	Design / evaluation	Long dsRNA / siRNA	siRNA length for <i>in silico</i> 'dicer' to identify potential off-target effects in the transcriptome	19 bp
Exclude regions of low sequence complexity	Design / evaluation	Long dsRNA / siRNA	Filter ('mdust') to avoid regions of low complexity, such as simple nucleotide repeats	Yes
Exclude >5× CA[ACGT] repeats	Design / evaluation	Long dsRNA / siRNA	Filter to avoid longer stretches of CAN repeats	Yes
Efficiency scoring method	Design / evaluation	Long dsRNA / siRNA	Method 'weighted' [Shah <i>et al.</i> (12)] or 'rational' [Reynolds <i>et al.</i> (11)]	Weighted
Minimal siRNA efficiency score	Design / evaluation	Long dsRNA / siRNA	Efficiency score filter for 'diced' siRNAs, values between 0 and 100	Long dsRNAs: 20, siRNAs: 70
siRNA seed-match analysis	Design / evaluation	Long dsRNA / siRNA	siRNA seed-matches to a user-defined sequence database (FASTA) are calculated	Disabled
Maximum overlap with introns allowed	Design	Long dsRNA	Adjust the maximal allowed overlap of designs to intron-regions (in percent of length of design)	25%
Allow relaxation of parameters if required	Design / evaluation	Long dsRNA / siRNA	Allows to design reagents even if specificity, efficiency and low-complexity criteria are not met by the input sequence	Yes
Map designs to the genome	Design / evaluation	Long dsRNA / siRNA	Mapping of calculated designs to the genome (required for GBrowse visualization)	Yes
Calculate homology of designs to transcripts	Design / evaluation	Long dsRNA / siRNA	Mapping of calculated designs to the transcriptome using Blast for homology evaluation (below the defined <i>E</i> -value)	Yes (long dsRNAs: 1e-5, siRNAs: 0.1)
Reagent source type	Evaluation	Long dsRNA / siRNA	Source ('genomic' or 'CDS') the reagent was designed and synthesized with	Genomic
Primer design options	Design / evaluation	Long dsRNAs	Define optimal/minimal/maximal primer size, amplicon size range, primer pair penalty cutoff, number of primer designs per 'favorable' region	Opt: 20 bp, Min: 18 bp, Max: 27 bp, Range: 150–500 bp, Penalty: 4, Designs: 50
Add promoter sequence or tag	Design	Long dsRNAs	Promoter sequence (for <i>in vitro</i> transcription) or individual tag sequence to be added 5' of each primer	None
Number of designs reported per query sequence	Design	Long dsRNA / siRNA	Number of best long dsRNAs / siRNAs for each input sequence reported in the output	1
GFF file output	Design / evaluation	Long dsRNA / siRNA	Output of generic feature file for GBrowse database upload and documentation of design location	No
AFF file output	Design / evaluation	Long dsRNA / siRNA	Output of annotation file format for direct upload and visualization of designs in any GBrowse installation	No
Design of independent RNAi reagents	Design	Long dsRNA / siRNA	Sequences of previous designs uploaded as FASTA file will be excluded from new designs, can also be used to exclude other sequences (e.g. miRNA sites)	No
Analysis of siRNA pools	Evaluation	siRNA	Summary report for siRNA pools, if a tab-delimited file with 'siRNAID' and 'POOLID' columns was provided	No

^aDetailed descriptions including examples are available in the E-RNAi Wiki at: <http://e-rnai.org/wiki>.

structures of siRNAs for the 'rational' method using the Vienna RNA package (23).

In case of siRNAs, the calculated efficiency score directly refers to the efficiency of the sequence. siRNAs that do not match the user-defined efficiency criteria are flagged during the design.

Long dsRNAs contain many different siRNAs and scores are aggregated during processing. E-RNAi reports two scores for the efficiency of long dsRNAs: an average efficiency score of all contained siRNAs and the absolute number of efficient siRNAs (efficiency above the default or user-defined cutoff).

Design 1: twi:1_1

dsRNA information

Primer forward		Amplicon sequence
Sequence	taatacgactcactatagggGTTAGCGATCTCAACGGAGG	GTTAGCGATCTCAACGGAGGAGTCATGTCGCCCCGCTGCTTGGCGGATGACGGCAGTGCCGGCAGTTTGCTGGACGGATCCGATGCCGGCGGAAAGGCCTTCCGCAAGCCACGTCGCCGGCTGAAGCGGAAGCCCAGCAAGACGGAGGAGACGGACGAGTTCAGCAACCAGCGGGTCATGGCCAATGTGAGGGAGCGCCA
Length [nt]	20	Amplicon length [nt]
Tm[°C]	59.836	248
GC[%]	55.000	Amplicon location
Primer reverse		2R:18934924..18935171(+)
Sequence	taatacgactcactatagggATGATCTGCTGCAGGGACTT	
Length [nt]	20	
Tm[°C]	59.834	
GC[%]	50.000	
Primer pair penalty		0.3301

Target information

Intended target gene	FBgn0003900
Intended target transcripts (hits)	FBtr0100130 (230), FBtr0071953 (230)
Other targeted gene(s)	NA
Other targeted transcripts (hits)	NA

Reagent quality

siRNAs [19 nt]	On-target	Off-target	No-target	Efficient siRNAs	Avg efficiency score	LowComplexRegions	CAN
230	230	0	0	209	43.23	0	0

Additional quality evaluation

Sequence homology (e-value)	FBgn0003900(9e-139)
------------------------------------	---------------------

Genome Browser

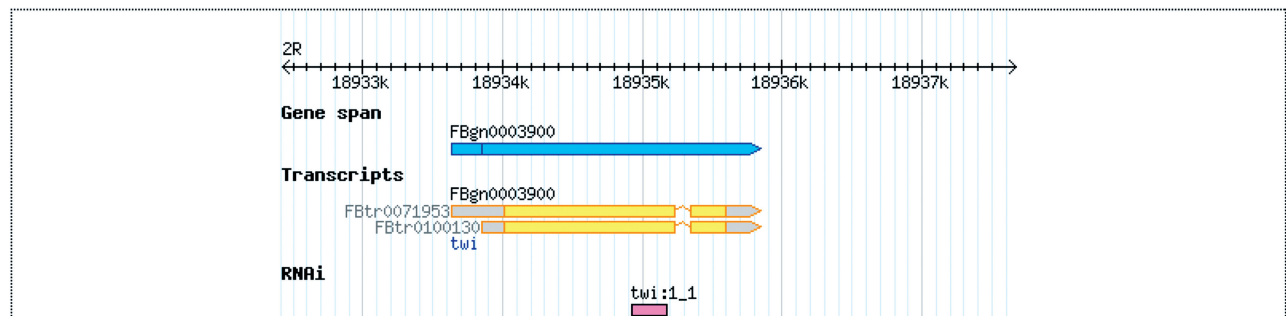


Figure 3. Report for the design of a long dsRNA against the gene *twi*. ‘dsRNA information’ summarizes the properties of the primers required to amplify the reagent from genomic or cDNA sources and also shows the full sequence of the construct, its length and location in the genome. ‘Target information’ lists all ‘intended’ and ‘other’ target genes found as well as all transcripts belonging to them (including the number of siRNA ‘hits’ to each transcript). ‘Reagent quality’ summarizes the analyzed quality parameters and shows the number of specific (‘on-target’) siRNAs contained within the dsRNA, the number of ‘off-target’ and ‘no-target’ siRNAs as well as the number of ‘efficient siRNAs’, the average siRNA efficiency and low-complexity information. In addition, all targets with significant ‘sequence homology’ to the long dsRNA are listed. GBrowse visualizes the location of the designed reagent in the gene model.

Amplicon design for long dsRNAs

The amplification of DNA templates for long dsRNAs by PCR from genomic or cDNA sources is a required step during the synthesis of long dsRNAs. E-RNAi uses primer3 (24) for the identification of suitable primers. The primer design can be influenced by user preferences (Table 2). We have found that the standardized primer design with primer3 facilitates similar PCR synthesis efficiency and the selection of smaller windows for the ‘amplicon size range’ (resulting in designs of comparable lengths) facilitates similar *in vitro* transcription reactions (data not shown). T7 or SP6 promoter sequences for *in vitro* transcription (or any other, individual tag) can be automatically added to the primer designs. The E-RNAi web service outputs a ready-to-use list for the synthesis of long dsRNAs.

Ranking of dsRNAi reagents

During the design of long dsRNAs, E-RNAi identifies suitable target sites by filtering the input sequences for regions that fulfill the default or user-defined criteria for predicted specificity, efficiency and complexity. These ‘favorable’ regions are used for automated primer design. If no suitable primer designs can be found, E-RNAi connects the closest, ‘favorable’ neighbors (if enabled by the user). This step is repeated until the design of primers was successful. In the final scoring of all possible dsRNAs, the ‘preferred’ reagent is determined by ranking all successful designs according to their predicted specificity and efficiency. Specificity of a long dsRNA is calculated as a percentage of siRNAs targeting the intended transcripts. Efficiency refers to the absolute number of efficient siRNAs (user-defined) contained in the long dsRNA.

The evaluation of RNAi reagents implements the same procedures as used for the design. E-RNAi reports e.g. how many siRNAs are contained in a long dsRNA predicted to be unspecific or how many siRNAs have an efficiency score above the user-defined cutoff.

As for the design of long dsRNAs, the design of siRNAs also implements filters to avoid regions that do not fulfill the user-defined criteria regarding predicted specificity, efficiency and complexity. siRNAs are finally ranked for specificity and efficiency.

E-RNAi for the design of independent RNAi reagents

Independent RNAi reagents are a necessary experimental step to confirm RNAi phenotypes. The updated version of E-RNAi offers several tools for their design. The user can upload sequences of pre-existing reagents (as a FASTA file) that will be excluded in the design process. Alternatively, independent target regions can be selected from sequence and gene-identifier queries or through the GBrowse interface (Figure 2).

E-RNAi output formats

E-RNAi results are presented as a webpage that can also be downloaded and locally viewed in any web browser. Linked pages provide detailed information for each design such as its sequence, primer and target information.

An example result page for RNAi reagents targeting the *Drosophila* gene *twi* is shown in Figure 3. Information of RNAi designs is also provided as tab-delimited files, FASTA sequence files as well as GFF and AFF files.

DISCUSSION

The design of specific and potent long dsRNAs and siRNAs is of key importance for RNAi experiments. The E-RNAi web service implements a user-friendly and straightforward selection of RNAi design templates and implements all steps required for the prediction of specific and efficient RNAi reagents. Moreover, available designed reagents can be evaluated and re-annotated using E-RNAi or used as sequence to be avoided for independent design.

In addition to classical model organisms such as *Caenorhabditis elegans* and *Drosophila melanogaster*, E-RNAi incorporates organisms that were only recently sequenced and annotated. Experiments using RNAi in these organisms can benefit from the knowledge about e.g. quality parameters of RNAi reagents obtained from studies in classical systems.

The update of E-RNAi also implements new findings with regards to RNAi specificity and efficiency, including analysis of sequences for low-complexity regions or improved efficiency predictors. We will continue to update the web service as new results become available. Technically, the new database structure behind the web server greatly facilitates the addition of organisms where genome and transcriptome sequences are available. The implementation of faster alignment tools (primarily Bowtie) significantly shortened the time to calculate designs and was also important to allow batch queries. The maximal size for batch queries will be further increased in the future, offering the possibility to e.g. re-annotate complete RNAi libraries online.

A comprehensive documentation of all E-RNAi input and output features as well as a detailed description of all settings is available in the E-RNAi Wiki at <http://www.e-rnai.org/wiki>, including tutorial sections e.g. about the calculation of independent designs.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We are grateful to Thomas Sandmann for critical comments on the manuscript.

FUNDING

Studienstiftung (to T.H., PhD fellowship); Deutsche Forschungsgemeinschaft and the Helmholtz Association (partial); EC FP7 Program (Grant 201666). Funding for open access charge: Intramural funding.

Conflict of interest statement. None declared.

REFERENCES

1. Boutros,M. and Ahringer,J. (2008) The art and design of genetic screens: RNA interference. *Nat. Rev. Genet.*, **9**, 554–566.
2. Fire,A., Xu,S., Montgomery,M.K., Kostas,S.A., Driver,S.E. and Mello,C.C. (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*, **391**, 806–811.
3. Chapman,E.J. and Carrington,J.C. (2007) Specialization and evolution of endogenous small RNA pathways. *Nat. Rev. Genet.*, **8**, 884–896.
4. Clemens,J.C., Worby,C.A., Simonson-Leff,N., Muda,M., Maehama,T., Hemmings,B.A. and Dixon,J.E. (2000) Use of double-stranded RNA interference in *Drosophila* cell lines to dissect signal transduction pathways. *Proc. Natl Acad. Sci. USA*, **97**, 6499–6503.
5. Bernstein,E., Caudy,A.A., Hammond,S.M. and Hannon,G.J. (2001) Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature*, **409**, 363–366.
6. Elbashir,S.M., Harborth,J., Lendeckel,W., Yalcin,A., Weber,K. and Tuschl,T. (2001) Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature*, **411**, 494–498.
7. Kulkarni,M.M., Booker,M., Silver,S.J., Friedman,A., Hong,P., Perrimon,N. and Mathey-Prevot,B. (2006) Evidence of off-target effects associated with long dsRNAs in *Drosophila melanogaster* cell-based assays. *Nat. Methods*, **3**, 833–838.
8. Anderson,E.M., Birmingham,A., Baskerville,S., Reynolds,A., Maksimova,E., Leake,D., Fedorov,Y., Karpilow,J. and Khvorova,A. (2008) Experimental validation of the importance of seed complement frequency to siRNA specificity. *RNA*, **14**, 853–861.
9. Birmingham,A., Anderson,E.M., Reynolds,A., Ilsley-Tyree,D., Leake,D., Fedorov,Y., Baskerville,S., Maksimova,E., Robinson,K., Karpilow,J. *et al.* (2006) 3' UTR seed matches, but not overall identity, are associated with RNAi off-targets. *Nat. Methods*, **3**, 199–204.
10. Ma,Y., Creanga,A., Lum,L. and Beachy,P.A. (2006) Prevalence of off-target effects in *Drosophila* RNA interference screens. *Nature*, **443**, 359–363.
11. Reynolds,A., Leake,D., Boese,Q., Scaringe,S., Marshall,W.S. and Khvorova,A. (2004) Rational siRNA design for RNA interference. *Nat Biotechnol.*, **22**, 326–330.
12. Shah,J.K., Garner,H.R., White,M.A., Shames,D.S. and Minna,J.D. (2007) siR: siRNA information resource, a web-based tool for siRNA sequence design and analysis and an open access siRNA database. *BMC Bioinformatics*, **8**, 178.
13. Wang,X., Varma,R.K., Beauchamp,L., Magdaleno,S. and Sendera,T.J. (2009) Selection of hyperfunctional siRNAs with improved potency and specificity. *Nucleic Acids Res.*, **37**, e152.
14. Echeverri,C.J., Beachy,P.A., Baum,B., Boutros,M., Buchholz,F., Chanda,S.K., Downward,J., Ellenberg,J., Fraser,A.G., Hacohen,N. *et al.* (2006) Minimizing the risk of reporting false positives in large-scale RNAi screens. *Nat. Methods*, **3**, 777–779.
15. Henschel,A., Buchholz,F. and Habermann,B. (2004) DEQOR: a web-based tool for the design and quality control of siRNAs. *Nucleic Acids Res.*, **32**, W113–W120.
16. Huesken,D., Lange,J., Mickanin,C., Weiler,J., Asselbergs,F., Warner,J., Meloon,B., Engel,S., Rosenberg,A., Cohen,D. *et al.* (2005) Design of a genome-wide siRNA library using an artificial neural network. *Nat. Biotechnol.*, **23**, 995–1001.
17. Tafer,H., Ameres,S.L., Obernosterer,G., Gebeshuber,C.A., Schroeder,R., Martinez,J. and Hofacker,I.L. (2008) The impact of target site accessibility on the design of effective siRNAs. *Nat. Biotechnol.*, **26**, 578–583.
18. Arziman,Z., Horn,T. and Boutros,M. (2005) E-RNAi: a web application to design optimized RNAi constructs. *Nucleic Acids Res.*, **33**, W582–W588.
19. Stein,L.D., Mungall,C., Shu,S., Caudy,M., Mangone,M., Day,A., Nickerson,E., Stajich,J.E., Harris,T.W., Arva,A. *et al.* (2002) The generic genome browser: a building block for a model organism system database. *Genome Res.*, **12**, 1599–1610.
20. Langmead,B., Trapnell,C., Pop,M. and Salzberg,S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
21. Kent,W.J. (2002) BLAT—the BLAST-like alignment tool. *Genome Res.*, **12**, 656–664.
22. Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
23. Hofacker,I.L. (2009) RNA secondary structure analysis using the Vienna RNA package. *Curr. Protoc. Bioinformatics*, Chapter 12:Unit12 12.
24. Rozen,S. and Skaletsky,H. (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.*, **132**, 365–386.